

Northern Star segmentation policy

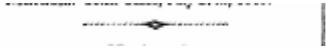
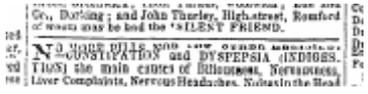
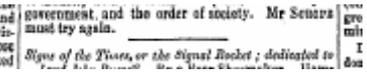
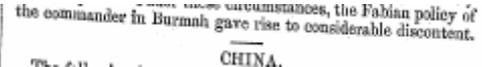
The Northern Star is a complex text as it has little consistent structure. However, it does correspond to our model of departments and items, and these can be identified through typographical rules.

0.0 General Points

- 0.1 In this title there is often more than one department on a page. Therefore it is essential that the sequencing problem – where items appear in the wrong order in the ToC – is addressed.
- 0.2 There may be a problem segmenting items where the page image is not square, perhaps due to tight binding of the hard copy. In this instance it is better to get the whole of an item in a segment – even if it means including parts of the segments that surround it – rather than missing out content.

1.0 Items

Items can be recognized by horizontal rules. These can be in a variety of formats including:

- 1.1 
- 1.2 
- 1.3 
- 1.4 
- 1.5 
- 1.6 

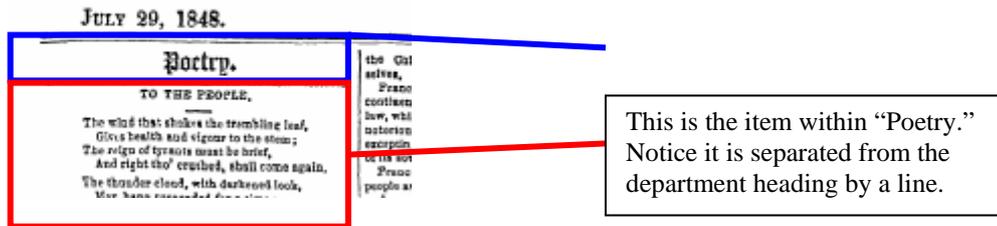
Just as in the *Leader* and *Publishers' Circular*, the advertisements complicate this rule as they often use horizontal lines within themselves. Advertisements are divided by a horizontal line all the way across the column as in 1.4 above.

The small line in 1.6 is used to mark the division between items. However, it is also sometimes used to underline a heading. When it appears under a heading it should be ignored.

2.0 Departments

- 2.1 Any heading in gothic script is a department header. For instance:

This is the department header and should appear in the ToC.

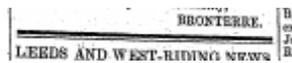


This is the item within “Poetry.”
Notice it is separated from the
department heading by a line.

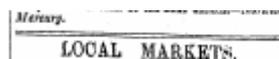
“Poetry” is a department header here, and should be segmented as shown so that it can appear in the ToC. The item is “To the People” and the verses that follow.

2.2 There are some departments that are not in gothic type. These can be identified by their appearance after double horizontal lines that run all the way across a column.

2.2.1 Sometimes they are double lines with one line darker than the other. For instance in 1838 they look like this:



2.2.2 Sometime they are simply double lines. For instance in 1842 they look like this:



2.3 You will notice that headers in gothic are preceded by double lines: this is because they are departments!

3.0 Advertisements

Advertisements are normally gathered together in clusters, often on the first and second pages and/or sometimes on the middle and last pages. Given that Olive operators can recognise adverts well we would if possible like them to create a department called ‘Advertisements’ (either with text or a snippet of the first advertisement) in the TOC whenever they see a cluster like this. The advertisement sections are preceded by double lines, and double lines mark where they end.

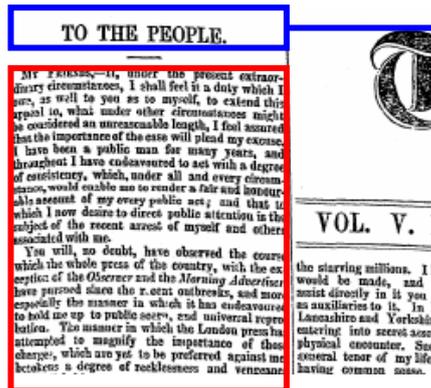
4.0 Problems

Although the *Northern Star* is the least structured of our titles, the segmentation policy is quite straightforward. Items are easy to identify as they are demarcated from each other with dividers. Departments are those items either in gothic or preceded by a double line (of some sort) that runs all the way across a column.

The only place where this is problematic is when an item starts at the top of a column. In this case there are no dividers so it might be difficult to tell whether an item is a department or not. Below we describe two departments that always appear but never in gothic type:

4.1 There are three departments of which you should be especially aware:

3.1.1 The first is that the first column on the first page is **usually** devoted to a leading article by the proprietor, Fergus O'Connor. For instance:



“To the People” is the department header and should appear as a snippet in the ToC. The red box marks the item.



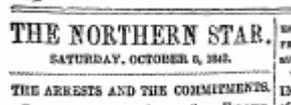
“To Readers and Friends” should be the department header, and appear in the ToC entry/snippet. The red boxes mark the items within the department.

The rest of the front page is taken up by advertisements. There are occasions where this department does not appear. It is easy to spot when this is the case (for instance for portions of 1838) as advertisements take its place.

4.1.2 The second is the regular leading article usually entitled “The Northern Star.” This is always a department, is usually on p. 3 or 4, but does not have gothic type, for instance:



4.1.2.1 “The Northern Star” 24 March 1838, p. 3.



4.1.2.2 “The Northern Star” 8 October 1842, p. 4.

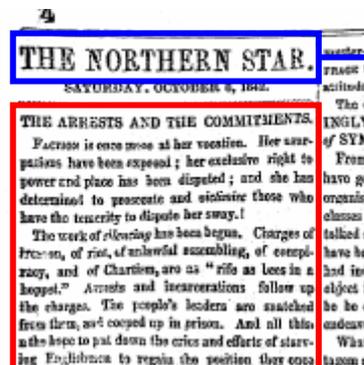


4.1.2.3 “The Northern Star” 29 July 1848, p. 4.



4.1.2.4 “The Star” 24 April 1852, p. 4. It is called “The Star” as the newspaper is now called the *Star of Freedom*.

In each case the department header is the portion between the two double lines, for instance:



“The Northern Star” is the department header, what follows (in red boxes) are the items.

4.1.3 “Markets.” There is always a report on the markets towards the end of each

number of the *Northern Star*. It is usually marked as a department with either double lines, gothic text, or both. However occasionally it isn't. For instance:



In this example, from 1852, “Markets” would not be identified as a department as the double line does not run all the way across the page. However, “Markets” should always appear in the ToC.

Jim Mussell and Suzanne Paylor